Let's use the normal approximation. Thus $Z = (X - np)/\sqrt{npq}$ is approximately a standard normal variable when n is large. In the current case, were interested in the probability of rejecting the claim, or $P(X < 75)$. Recalling that we are approximating the area under the histogram whose bars are centered on the integers, this is $n = 200$ $p = .4$, $q = 1 - p = .6$ so

$$P\left(Z \le \frac{X - np}{\sqrt{npq}}\right) = P\left(\frac{74.5 - (200)(.4)}{\sqrt{200(.4)(.6)}}\right) = P(Z \le -.794) = .2160$$

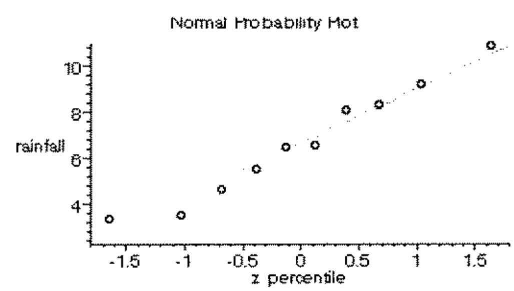
interpolating Table A.3 on P. 740.

According to the rule of thumb, this approximation is valid if $np \ge 10$ and $nq \ge 10$. In the present case, $nq \ge np = (200)(.4) = 80 \ge 10$ and the approximation is valid.

(14.) *Suppose the following annual rainfall readings in inches were observed at Dead Horse Point for the years 1992 to 2001. Construct a normal probability plot for these readings. Is normality plausible?*

6.47, 8.07, 8.34, 9.21, 3.54, 3.39, 10.91, 6.54, 4.61, 5.51

The normal probability plot is obtained as follows. We sort the data so $y_1 = 3.39, \ldots, y_{10} = 10.91$. We find the critical normal values $\Phi(x_i) = (i - 0.5)/n$ so that $x_1 = z_{.95} = -1.645$ and so on. Then plot the points $[x_i, y_i]$. Here is the complete list: $[-1.645, 3.39]$, $[-1.036, 3.54]$, $[-.674, 4.61]$, $[-.385, 5.51]$, $[-.126, 6.47]$, $[.126, 6.54]$, $[.385, 8.07]$, $[.674, 8.34]$, $[1.036, 9.21]$, $[1.645, 10.91]$.



The data lines up quite well so normality of the data is plausible. (In fact a normal random number generator generated the data.)

(15.)*2% of the electric toasters made by the Tremonton Toaster Company will require repairs within 90 days after they are sold. Determine the probability that among 1200 toasters sold, at least 30 will require repairs within the first 90 days after they are sold. Use an approximation. Why is your approximation justified?*

Let X be the number of toasters needing repair. We may use the normal approximation of binomial because the rule of thumb is satisfied: $nq \ge np = 1200 \times 0.02 = 24 \ge 10$. The desired probability is $P(X \ge 30) = 1 - P(X \le 29) = 1 - B(29, 1200, 0.02) \approx$

$$1 - \Phi\left(\frac{X + 0.5 - np}{\sqrt{npq}}\right) = 1 - \Phi\left(\frac{29.5 - 24}{\sqrt{(1200)(0.02)(0.98)}}\right) = 1 - \Phi(1.134) = \Phi(-1.134) = \boxed{0.1300}.$$

(16.) Lifetimes of a certain component are lognormally distributed with parameters $\mu = 1.000$ and $\sigma = 0.500$. Find the mean lifetime of these components. Find the standard deviation of the lifetimes. Find the probability that the component lasts longer than four days.

Let X represent the lifetime of a randomly selected component. The mean of X is found by the formula $\mu = \exp(\mu + \frac{1}{2}\sigma^2) = \exp(1 + \frac{1}{2}(0.5)^2) = 3.08$ days. The variance is $V(X) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1] = \exp(2 \cdot 1 + (0.5)^2)[\exp((0.5)^2) - 1] = 2.6948$. The standard deviation is therefore $\sigma = \sqrt{V(X)} = \sqrt{2.6948} = 1.64$ days. Let $Y = \ln(X)$ be the logarithm of the variable. Since X is lognormal, $Y \sim N(1, 0.5)$ is normally distributed, thus $Z = \dfrac{Y - \mu}{\sigma}$ is a standard normal variable. Thus $P(X > 4) = 1 - P(X \le 4) = 1 - P(\ln X \le \ln 4) = 1 - P\left(Z \le \dfrac{\ln 4 - 1}{0.5}\right) = 1 - \Phi\left(\dfrac{1.386 - 1.000}{0.500}\right) = 1 - \Phi(.772) = 1 - .7800 = .2200$. We have interpolated. The table gives $\Phi(.77) = .7794$ and $\Phi(.78) = .7823$ so $\Phi(.772) = .7794 + \dfrac{.002}{.01}(.7823 - .7794) = .7800$.

(hypergeometric) may be approximated by sampling with replacement (binomial.) Here $n/N = 3/120 = .025$ so, yes, it is an acceptable approximation.

(7.) *Suppose 35% of all computer scientists have worked on a neural-net program. A representative for a company seeking computer scientists who have worked on neural-net programs will continue to interview computer scientists until she finds two who have worked on a neural-net program. What is the probability that the interviewing process will end with the fifth interview? On or before the fifth interview? What is the average number and standard deviation of the number of interviews that need to be done before she finds two computer scientists who have worked on neural-net programs?*

This is an example of the negative binomial distribution. Each interview has a $p = .35$ chance of success (finding someone who has worked on a neural-net program.) The independent interviews continue until a total of $r = 2$ successes have been obtained. X is the number of failures that precede the r-th success. The process ends on the fifth interview ($x = 3$ failures and $r = 2$ successes)

$P(X = 3) = \text{nb}(3; 2, .35) = \binom{4}{1}(.35)^2(.65)^3 = .135.$

The probability that it ends on or before the fifth interview is

$P(X \le 3) = \text{nb}(0; 2, .35) + \text{nb}(1; 2, .35) + \text{nb}(2; 2, .35) + \text{nb}(3; 2, .35)$
$= \binom{1}{1}(.35)^2(.65)^0 + \binom{2}{1}(.35)^2(.65)^1 + \binom{3}{1}(.35)^2(.65)^2 + \binom{4}{1}(.35)^2(.65)^3 = .572$

Using the expected value and standard deviation for a negative binomial, the average number of interviews $I = 2 + X$ is

$E(2 + X) = 2 + \frac{r(1-p)}{p} = 2 + \frac{2(.65)}{(.35)} = 5.714,\ \sigma_I = \sigma_X = \frac{\sqrt{r(1-p)}}{p} = \frac{\sqrt{2(.65)}}{.35} = 3.258.$

(8.) *Suppose that the number of parasites that are infecting a host has a Poisson Distribution with a mean of 10. What is the probability that there are 8 parasites on the host? There are 8 or 9 parasites on the host? There are no parasites?*

The average of a Poisson distribution gives the parameter $\lambda = 10$. Thus the probabilities of $X = 8$, $8 \le X \le 9$ and $X = 0$ are

$p(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$; $p(8; 10) = \frac{e^{-10}10^8}{8!} = .113,$

$P(8 \le X \le 9) = p(8; 10) + p(9; 10) = \frac{e^{-10}10^8}{8!} + \frac{e^{-10}10^9}{9!} = .238,\ P(X = 0) = p(0, 10) = \frac{e^{-10}10^0}{0!} = .0000454.$

Alternatively, one could use the cumulative Poisson values tabulated on p. 739.

$p(8, 10) = P(8, 10) - P(7, 10) = .333 - .220 = .113;\ P(X = 0) = p(0, 10) = P(0, 10) = .000.$

$P(8 \le X \le 9) = P(9, 10) - P(7, 10) = .458 - .220 = .238.$

(9.) *Suppose a continuous random variable has the distribution function of the form $f(x) = c\sin x$ for $0 \le x \le \pi$, and $f(x) = 0$ otherwise. Find the constant c. Find the cumulative distribution function. Find the mean, median and standard deviations. What is the probability that $X > \pi/4$?*

The distribution function of a continuous random variable has to satisfy $f(x) \ge 0$ and have total integral equal to one. Calfulating the integral, we need

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\pi} c\sin x\,dx = c\Big[-\cos x\Big]_0^{\pi} = c\Big[-(-1) + (1)\Big] = 2c,$$

so $c = .5$. The cumulative distribution function is

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \begin{cases} 0, & \text{if } x < 0; \\ \int_0^x \frac{\sin x\,dx}{2} = \frac{1-\cos x}{2}, & \text{if } 0 \le x \le \pi; \\ 1, & \text{if } 1 < x. \end{cases}$$

The mean and expected square are (finding the antiderivatives by trial and error=guessing)

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^{\pi} \frac{x\sin x\,dx}{2} = \left[\frac{\sin x - x\cos x}{2}\right]_0^{\pi} = \frac{\pi}{2};$$
$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^{\pi} \frac{x^2\sin x\,dx}{2} = \left[\cos x + x\sin x - \frac{x^2\cos x}{2}\right]_0^{\pi} = \frac{\pi^2}{2} - 2.$$

Since the distribution is symmetric about $x = \pi/2$, $(f(\pi/2 + x) = f(\pi/2 - x))$ the median (and the mean) is the center point $\mu = \tilde{\mu} = \pi/2$. To check, $P(X \le \pi/2) = F(\pi/2) = 1/2$. The short cut formula gives the standard deviation $\sigma = .684$, since

$$\sigma^2 = E(X^2) - [E(X)]^2 = \frac{\pi^2}{2} - 2 - \frac{\pi^2}{4} = \frac{\pi^2}{4} - 2.$$

that a random concrete sample have a compressive strength larger that 6228 psi? (b.) Below what psi will 20% of the samples of concrete be? [Answer to three decimal places.]

Let X denote the rv compressive strength, which is normally distributed with $\mu = 6000.$ and $\sigma = 240.0$ psi. Then, standardizing, $P(X > 6228) = P\left(Z = \frac{X-\mu}{\sigma} > \frac{6228-60000}{240.0}\right) = P(Z > 0.9500) = P(Z < -.9500) = \boxed{0.1711}$. The critical value $z_{0.8}$ we desire satisfies $P(Z < z_{0.8}) = 0.200$. The table on p. 740 has $\Phi(-0.840) = .2005$ and $\Phi(-0.850) = .1977$. Interpolating $\dfrac{-0.840 - z_{0.8}}{-0.840 + 0.850} = \dfrac{.2005 - .2000}{.2005 - .1977}$ so $z_{0.8} = \boxed{-0.842}$. Thus the critical $X = z_{0.8}\sigma + \mu = (-0.842)(240.) + 6000. = \boxed{5.80 \times 10^3}$.

(5.) *The Castle Dale Candy Company claims that at most five of a shipment of twentyfive chocolate turkeys have a cream-filled center. To test this claim, four turkeys will be selected at random from the shipment and tasted. The claim is to be rejected if two or more turkeys are found to have a cream-filled center. What is the probability that the claim is rejected, given that the shipment actually contained five cream-filled turkeys?*

The number of cream-filled turkeys, X, in $n = 4$ trials is a hypergeometric variable. There are M "successes" or "cream-filled" in a total of $N = 25$. Thus the probability of the claim being rejected when there are $M = 5$ cream-filled is $P(X \geq 2) = 1 - P(X \leq 1) = 1 - \text{Hyp}(1, 4, 25, 5) = 1 - \text{hyp}(0, 4, 25, 5) - \text{hyp}(1, 4, 25, 5) =$

$$P(X \geq 2) = 1 - \frac{\binom{5}{0}\binom{20}{4}}{\binom{25}{4}} - \frac{\binom{5}{1}\binom{20}{3}}{\binom{25}{4}} = 1 - \frac{1 \cdot 4845}{12650} - \frac{5 \cdot 1140}{12650} = \boxed{0.166}$$

More Practice Problems.

(2.) *Suppose three cards are randomly selected from a standard deck of 52 cards without replacement. Let A_1 denote the event that three clubs are drawn. Let A_2 denote the event that a king, a queen and a jack are drawn. Are these independent events? What if the cards were drawn with replacement?*

First consider the case without replacement. The event that three clubs are drawn without replacement is a hypergeometric random variable $x = 3$ the number of successes (clubs) drawn from an urn with $N = 52$ cards of which $M = 13$ are successes out of a total of $n = 3$ cards drawn. There are four ways to choose a king, four to choose a queen and four to choose a jack, but only one way to do this given if all are clubs. Thus

$$P(A_1) = h(3; 3, 52, 13) = \frac{\binom{13}{3}\binom{39}{0}}{\binom{52}{3}} = \frac{286 \cdot 1}{22{,}100} = .0129,$$

$$P(A_2) = \frac{4 \cdot 4 \cdot 4}{\binom{52}{3}} = \frac{64}{22{,}100} = .00290. \qquad P(A_1 \cap A_2) = \frac{1}{\binom{52}{3}} = \frac{1}{22{,}100} = .0000452.$$

Since $P(A_1)P(A_2) = .0000374 \neq P(A_1 \cap A_2)$ these events are not independent.

Next consider the case with replacement. This time the probability of ♣ is $p = .25$ on each draw. This becomes a binomial with $x = 3$ successes in $n = 3$ trials. There are still four ways to choose a king, four to choose a queen and four to choose a jack, but still only one way to do this given if all are clubs. Remember to divide out 3! since order is not counted.

$$P(A_1) = \text{bin}(3; 3, .25) = \binom{3}{3}(.25)^3(.75)^0 = \frac{1}{64} = .0156$$

$$P(A_2) = \frac{4 \cdot 4 \cdot 4}{\frac{52^3}{3!}} = .00273, \qquad P(A_1 \cap A_2) = \frac{1}{\frac{52^3}{3!}} = .0000427.$$

This time $P(A_1)P(A_2) = P(A_1 \cap A_2)$ so these events are independent.

(3.) *Suppose a random variable takes the values $D = \{1, 3, 5, 7, 9\}$. Suppose that its probability mass function is given by $p(1) = .2$, $p(3) = .3$, $p(5) = .2$, $p(7) = .1$, $p(9) = .2$ and $p(x) = 0$ if $x \notin D$. Find the cumulative distribution function. Sketch graphs of the pmf and cdf. What is the probability that $2 \leq X \leq 7$? Find the expected values $E(X)$ and $E(X^2)$. What is the standard deviation? Suppose that $R(x) = 8 - x/2$. What is $E(R)$?*

Solutions of the Spring '01 Midterm Problems

(1.) *A Cache Valley cheese manufacturer claims that no more than 5% of all their packages contain less cheese than indicated on the label. To test this claim, 25 packages are randomly selected and weighed. The claim is accepted if fewer than 3 of the packages contain less cheese than indicated on the label. What is the probability that the claim is accepted if the actual percentage of packages with less cheese than indicated is 5%? 20%?*

This is a binomial random variable with $n = 25$ trials and with chance of success (finding that the package is underweight) equal to $p = .05$ and $p = .20$. The probability that the number of underweight packages X is fewer than 3 is $P(X \le 2) = \text{bin}(0, 25, p) + \text{bin}(1, 25, p) + \text{bin}(2, 25, p) = \text{Bin}(2; 25, p)$. This is tabulated on p.738. $\text{Bin}(2; 25, .05) = .873$ and $\text{Bin}(2; 25, .2) = .098$.

(2.) *Suppose that the continuous random variable X has a probability distribution function of the form $f(x) = 0$ if $x < 1$; $1/x$, if $1 \le x \le c$; and 0, if $c < x$. (a.) Find the constant $c > 1$ which makes $f(x)$ a probability distribution. (b.) Find the cumulative distribution function of X. (c.) Find the probability that $c^{\frac{1}{3}} < X < c^{\frac{1}{2}}$. (d.) Find median of X (50th percentile.) (e.) Find mean of X. (f.) Find the variance of X.*

To be a probability distribution, the total probability must be one. Thus

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_1^c \frac{dx}{x} = \ln(c) - \ln(1) = \ln(c)$$

so $c = e$. The cumulative distribution function is

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \begin{cases} 0, & \text{if } x < 1; \\ \int_1^x \frac{dx}{x} = \ln(x), & \text{if } 1 \le x \le e; \\ 1, & \text{if } e < x. \end{cases}$$

$$P(e^{1/3} < X < e^{1/2}) = \int_{e^{1/3}}^{e^{1/2}} \frac{dx}{x} = \ln(e^{1/2}) - \ln(e^{1/3}) = 1/2 - 1/3 = 1/6.$$

To find the median, solve for $\tilde{\mu}$ in $P(X \le \tilde{\mu}) = .50$ or $F(\tilde{\mu}) = \ln \tilde{\mu} = .5$ which implies $\tilde{\mu} = e^{1/2}$.

The mean is

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_1^e dx = e - 1 = 1.718.$$

The expected square is

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_1^e x\,dx = \left[\frac{x^2}{2}\right]_1^e = \frac{e^2-1}{2} = 3.195.$$

The short cut formula gives

$$V(X) = E(X^2) - [E(X)]^2 = \frac{e^2-1}{2} - (e-1)^2 = (e-1)\left(\frac{3-e}{2}\right) = .242.$$

(3.) *Suppose field mice are distributed at random in Juab County according to a Poisson Distribution with parameter $\alpha = 10$ per acre. (a.) What is the probability that there are at least six (6) but not more than nine (9) mice on a random acre? (b.) What is the expected number of mice that one would find on a random two (2) acre plot? (c.) How big should my sampling region be in order to be 90% sure of finding five (5) or more mice?*

For one acre, this is a Poisson variable with parameter $\lambda = \alpha = 10$ where X is the number of mice found on a random acre. Then $P(6 \le X \le 9) = \text{Poisson}(9; 10) - \text{Poisson}(5; 10) = .458 - .067 = .391$. (p. 739)

The parameter $\lambda = \alpha t$ is proportional to the number of acres t. The Poisson probability of k success is $p_k(t) = e^{-\alpha t}(\alpha t)^k/k!$. Thus when there are $t = 2$ acres, the parameter $\lambda = \alpha t = (10)(2) = 20$ and the expected number of mice is $E(X) = \lambda = 20$.

The last problem is to find the number of acres t so that $P(X \ge 5) = 1 - P(X \le 4) = .90$. In other words, for which $\lambda = \alpha t$ is it true that the cfd $\text{Poisson}(4; \lambda) = .10$? Looking at (p.739), one finds $\text{Poisson}(4; 8.0) = .100$. Thus $\alpha t = 10t = 8.0$ so $t = .8$ acres.

(4.) *A student commutes daily from his Murray home to The "U". The average time for a one way trip is 24 minutes with a standard deviation of 3.8 minutes. Assume that the duration of the trip is normally distributed.*